

ESTIMATING MEAN CONCENTRATIONS WHEN
SOME DATA ARE BELOW THE DETECTION LIMIT

BY

R. H. Shumway
A. S. Azari

*Division of Statistics
University of California, Davis
Davis, CA 95616*

Final Report
Contract ARB-A733-045
November 8, 1988

Submitted to:

Research Division
California Air Resources Board
P. O. Box 2815
Sacramento CA 95812

The statements and opinions in this report are those of
the authors and not necessarily those of the California
Air Resources Board.

TD 890

S485
1989

TABLE OF CONTENTS

1. Introduction
2. The Use of Transformations
3. Estimation of the Mean
4. Confidence Intervals
 - 4.1 The Delta Method
 - 4.2 Bootstrap Method
5. Simulation and an Example
6. Recommendations
7. Acknowledgements
8. References

APPENDIX A: Technical Details

APPENDIX B: Computer Programs

- B1. Introduction
- B2. Documentation
- B3. BASIC Program
- B4. FORTRAN Program

1. Introduction

The problems involved in monitoring levels of various substances in the environment for possible classification as toxic air contaminants is of public health concern because of the possible carcinogenic effects that such substances might have on man. In general, dose is considered to be the best measure of carcinogenicity and the mean concentration of a given substance can be used as a direct estimator for dose.

The true mean of the given substance is not known and must be estimated from a sample of observations collected from some environmental area of concern. Since measurements of toxic air contaminants involve potentially expensive sampling procedures and even more expensive laboratory determinations, the sample sizes used to characterize the mean level of a toxic contaminant are typically rather small. Even so, if the underlying data from which the sample has been drawn can be assumed to follow a normal distribution, the simple average of a sample of observations forms a satisfactory estimator for the mean. (Normally distributed data follows the familiar symmetric bell-shaped curve with no preponderance of values at either very high or very low levels.) The uncertainty of the sample mean as an estimator for the true mean is often summarized by presenting a confidence interval for the mean which gives a range of values that one would expect to cover the mean a given percentage (say 90%) of the time in repeated experiments.

It is unfortunate that for most environmental data, the assumption that the underlying distribution is normal will not be appropriate so that the usual sample mean will not be a good estimator for the population mean. One reason is that measurements made on toxic pollutants tend to have more observations than one would expect under normality, both at very

high levels and at very low levels. The analysis of non-normal data can be approached by attempting to find a transformation that produces a normally distributed population when applied to the data sample. While the estimation of the mean in the new scale can then be approached in the usual way using normal theory, a method must be found for transforming the estimators for the mean in the transformed scale back to an estimator for the mean in the original scale. The logarithmic transformation is often applied to pollution data leading to an estimation procedure based on the lognormal distribution. (see Gilliom and Helsel (1986), Ringdal (1975) and Blandford and Shumway (1982). In this report, we will consider the more general class of power transformations due to Box and Cox (1964) (see Johnson and Wichern (1988), pp. 155-162 for a more recent description).

A further complication is introduced by the fact that laboratory determinations can only be measured above some detection limit, making it impossible to know the exact sample values of observations collected that have apparent concentrations below this limit. Even if the data are normally distributed, this censoring means that it is now impossible to calculate the conventional sample mean since we have no way of deciding what to fill in for those values that are only measured as being below the detection limit. The problem of estimating the mean for censored data using the method of maximum likelihood (which yields the ordinary sample mean in the uncensored case) was first considered by Cohen (1959) and later by Ringdal (1975). We will use maximum likelihood estimators in this report because they provide a convenient means for handling the problems introduced with censored transformed data.

To summarize, in this report we seek a method for estimating the mean using small samples of non-normal environmental data that are subject to

censoring because of detection limits. We will also develop procedures for computing confidence limits for the mean using large-sample theory for maximum likelihood estimators (the delta method (Cramer (1946)) and a simulation procedure called the bootstrap (Efron (1979),(1982))). These techniques are described in Sections 2,3 and 4. An evaluation using a large range of contrived data is presented in Section 5.

2. The Use of Transformations

When environmental data cannot be modeled in terms of the normal distribution, the sample mean is not the appropriate estimator for the population mean. In such cases it may be appropriate to search for a function of the data values that conforms more closely to the normal distribution. Although the use of the logarithmic transformation for purposes of stabilizing variances and transforming to normality is well established in the literature, there may be occasions in which other transformations or perhaps even no transformation at all may be more appropriate.

A more general class of power transformations due to Box and Cox (1964) (see Johnson and Wichern (1988), pp. 155-162)) can be applied to the data and includes the logarithm, no transformation and various other power laws as special cases. Since environmental data are non-negative, the power transformations considered in this report are limited to (1) no transformation, (2) fourth root, (3) square root and (4) logarithmic. Using one of the transformations in this family leads one hopefully to a set of transformed data which follows a normal distribution. There are two problems that occur when the use of a transformation is considered.

The first problem is in deciding which of the power laws produces data that conforms best to the normal distribution. Box and Cox (1964) proposed evaluating the likelihood function under each of the proposed transformations assuming that the likelihood function of the transformed data is that of a set of normally distributed observations. Then, one simply chooses the transformation that produces the largest value for the likelihood function. This procedure has the virtue that it can be done automatically using a computer but this does not necessarily mean that it is the best method for choosing a transformation. The use of probability plotting and comparing to the ordinates of the normal distribution function is still highly recommended. It is also important to establish for any given type of experimental data some common transformation that produces normally distributed observations in the majority of cases. Later on we shall see how the confidence intervals computed are affected by choosing the wrong transformation.

The second problem that one has when carrying out an analysis on transformed data is that the mean of the transformed data is rarely the parameter of interest. While the transformed scale can be of great interest in some fields, (eg. the use of the logarithmic scale in measuring seismic magnitudes) the primary objective here is to obtain an estimate of the mean and a confidence interval in the original scale of measurement. The theoretical mean in the original scale will be a non-linear function of the means and variances in the transformed scale. Fortunately, we are again saved by the likelihood function since it can be maximized over means and variances in the transformed scale using normal distribution theory. Since maximum likelihood estimators of functions are the corresponding functions of the maximum likelihood estimators, we

obtain easily the maximum likelihood estimators for the mean in the original scale by simply taking the appropriate inverse functions of the estimators in the transformed scale.

3. Estimation of the Mean

We continue the discussion begun in the last section relating to estimating the mean, first in the transformed scale taking into account the censoring and then in the original scale. When there are observations that are only known to have been below some detection limit, these observations are said to be censored. Various procedures can be considered for estimating the mean of a set of observations when some of the observations are censored. For example, a simulation study using small normal data sets was performed by Gleit (1985) who compared the performance of numerous estimations, including several fill-in options involving the proxies zero, the threshold and the conditional expectation given the censoring. He concludes that the procedure of filling in conditional expectations leads to an estimator that has the smallest mean-square error of those considered.

The approach taken in this report will be to use the maximum likelihood procedure to estimate the mean and variance parameters in the transformed scale under censoring and then to use the properties of maximum likelihood estimators to get the maximum likelihood estimators in the original scale. One problem with this approach is that the likelihood function for the censored data involves the cumulative normal distribution function, which produces a non-linear likelihood that cannot be solved directly for the estimators in the transformed scale. An early treatment of this problem using maximum likelihood is Poirer (1976).

Complicated likelihood functions involving missing or incompletely observed data can be maximized using the Expectation Maximization (EM) Algorithm of Dempster et al (1978). Aitkin (1981) and Blandford and Shumway (1982) have applied this algorithm to the regression case. Basically, the approach makes use of the likelihood of the complete data under the assumption that the censored data points have been observed. One can then calculate the conditional expectation of this likelihood given the pattern of censoring. The EM Algorithm asserts that iteratively maximizing this restricted likelihood with respect to the mean and variance leads to a sequence of estimators that always increase the likelihood function of the original censored data sample. Furthermore, the sequence converges, in this case, to the unique maximizers of the likelihood function for any fixed value of the transformation parameter. Details and equations can be found in Shumway et al (1988) or Appendix B of this report.

The procedure of the preceding paragraph leads to estimators for the transformation power and for the mean and variance in the transformed scale. The technique discussed at the end of the previous section leads to estimators for the mean in the original scale. Again, details and equations can be found in Shumway et al (1988) or in Appendix A.

4. Confidence Intervals

The procedure described in the previous section leads to maximum likelihood estimators for the means in the original scale but does not immediately produce an estimator for the variance of the estimated mean or a confidence interval. With a confidence interval, one can make an assessment of the probable range within which the true mean can be

expected to lie. More precisely, a confidence interval is a range of values for the mean which might be expected to cover the mean a certain percentage of the time in repeated applications of the confidence interval methodology. For example, 90% confidence intervals should cover the true mean concentration 90% of the time over the long run. One should be careful not to interpret this as being a statement about the probability of the mean lying within the interval at any given trial which would be either 1 or 0.

Again, in developing confidence intervals we are faced with choices. One can note that the maximum likelihood estimators developed above will have a limiting normal distribution with a predictable mean and variance in large samples. This leads to the use of the delta method, usually credited to Cramer (1946). If we are not comfortable with large-sample theory because our samples are generally not large, a resampling method due to Efron (1979),(1982),1985),(1987) called the bootstrap may be useful. The adaptation of these two methods to the problems at hand are described below.

4.1 The Delta Method

The computation of the large-sample variance covariance matrix of the mean and variance in the transformed scale depends on technical manipulations to compute the second derivatives of the log-likelihood function and is given in Shumway et al (1988) and in Appendix A of this report. The computation of the large-sample variance of the mean in the original scale depends on expanding it in a Taylors Series about the mean and variance in the transformed scale and then using the central limit

result of Cramer (1946) for functions of asymptotically normal variables. The resulting large sample variances for the non-linear functions implied by using the four transformations mentioned earlier are presented in Shumway et al (1988) and in Appendix A of this report.

The manipulations of the preceding paragraph result in a large-sample expression for the variance of the maximum likelihood estimator. An approximate confidence interval for the mean can then be computed by substituting its large-sample standard deviation into the familiar formula for confidence intervals for estimated mean.

4.2 Bootstrap Methods

The bootstrap of Efron (1979) is a resampling method that develops confidence intervals for the mean from the sampling distribution of means estimated from the resampled data. To illustrate, suppose that we have a sample of 20 observations on a toxic contaminant and that several of the observations have only been measured as being below some fixed detection limit. The method of Section 3 is used to compute a maximum likelihood estimator for the mean in the original scale and we wish to have an idea of the basic uncertainty in this estimator.

Consider drawing a sample of 20 observations with replacement from the original sample in the previous paragraph. Since the sample is drawn with replacement so that the same element can appear more than once, the "bootstrap" sample drawn will almost certainly be different than the original sample and will yield a different estimator for the mean. This procedure can be repeated a large number of times and each time one will, with high probability, obtain a different sample and a different estimator for the mean. All of these maximum likelihood estimators for the mean can

be combined into an empirical cumulative distribution function by arranging them in ascending order from smallest to largest. Within this sample, one can compute values (called percentiles) which are such that 5% and 95% of the means lie respectively below the values. These percentiles define the 90% bootstrap confidence interval.

One can also use the bootstrap sample to calculate standard errors and to assess the bias problem with maximum likelihood estimators in small samples. Efron (1985) has proposed a procedure leading to a bias-corrected bootstrap interval which adjusts the bootstrap percentile distribution according to the location of the maximum likelihood estimator in that distribution. One can also standardize by adjusting the bootstrap estimator. The adjustment is made by subtracting the maximum likelihood estimator of the original sample from it and then dividing by the standard deviation of the original sample computed from the delta method leading to bootstrap t- intervals (see Efron (1982)).

5. Simulations and an Example

In order to be able to choose a method to recommend from those presented in Sections 2-4, a simulation study was designed using the following four files of samples as inputs.

- (a) 400 samples of size 20 with 10% censoring
- (b) 400 samples of size 20 with 20% censoring
- (c) 400 samples of size 50 with 10% censoring
- (d) 400 samples of size 50 with 20% censoring

The above samples were generated for (1) Normally distributed data, (2) Data for which the square root was normal and (3) Lognormally distributed data, making a total of 12 files. To ensure comparability all files were started with the same random number seed.

Table 1 shows how well the Box-Cox method for choosing the best power transformation performed on each of the samples. We note that the method

Table 1 Simulation results for censored mean estimation using the Box-Cox transformation: the proportion of times each transformation was chosen.

Correct Transf. to Normality	Sample Size	Cens.	Proportion of Times Each Transformation Chosen			
			None	Square Root	Fourth Root	log
None	20	10%	.70	.12	.05	.14
	20	20%	.70	.09	.04	.17
	50	10%	.70	.14	.09	.07
	50	20%	.70	.14	.05	.11
Square Root	20	10%	.50	.19	.13	.19
	20	20%	.55	.15	.09	.21
	50	10%	.37	.29	.18	.16
	50	20%	.42	.27	.16	.16
log	20	10%	.02	.13	.32	.53
	20	50%	.07	.21	.24	.48
	50	10%	.00	.03	.29	.68
	50	20%	.00	.08	.29	.63

works quite well when choosing no transformation (correct 70% of the time) or in choosing the logarithmic transformation (correct 48-68% of the time). Note that in the case of the logarithmic transformation, the method worked substantially better for the large samples. The differences in performances between 10% and 20% censoring were minor. The method has considerable difficulty in identifying the square root transformation when it is the correct transformation to normality, with the preponderance of mistakes slanted toward making no transformation.

Table 2 shows the converge of confidence intervals computed by four methods when the power transformation is chosen by the Box-Cox method. Since the nominal level for these intervals was chosen to be 90%, one can evaluate how well the intervals are covering the known true value by comparing the tabular entry to .90. Of course, the lengths of the intervals, given below the proportion, are important as well; given that the coverages are equal, one would prefer the shorter interval.

The delta method was the best performer overall with coverages of .89- .90 for the normal and square root normal data and .84-.86 for the lognormal data. The bias-corrected percentile interval (based on 1000 bootstrap replications) was the next best performer, but its performance seemed to suffer some degradation as the censoring increased from 10% to 20%. The other methods did less well, particularly on the lognormal data. The relatively less successful performance of the bootstrap based percentile methods is not a complete surprise; it has been studied by Loh (1987).

Table 2 Simulation results: coverage proportions of confidence intervals (nominal 90% coverage) for estimates of censored means computed using the Box-Cox transformation. The numbers in parentheses are the average lengths.

Correct Transf. to Normality	Sample Size	Cens.	Delta Method	Perc.	BC Perc.	Bstrp t
None	20	10%	.89 (.72)	.83 (.65)	.87 (.66)	.87 (.76)
	20	20%	.89 (.73)	.76 (.56)	.83 (.59)	.83 (.72)
	50	10%	.90 (.46)	.82 (.42)	.86 (.42)	.85 (.46)
	50	20%	.88 (.47)	.72 (.37)	.76 (.37)	.78 (.44)
Square Root	20	10%	.89 (1.84)	.82 (1.67)	.89 (1.72)	.87 (2.02)
	20	20%	.89 (1.92)	.75 (1.49)	.84 (1.58)	.79 (1.99)
	50	10%	.89 (1.19)	.82 (1.08)	.88 (1.10)	.84 (1.23)
	50	20%	.88 (1.22)	.71 (0.97)	.77 (1.00)	.73 (1.22)
log	20	10%	.84 (26.8)	.78 (25.4)	.86 (30.6)	.72 (44.5)
	20	20%	.84 (26.9)	.70 (23.1)	.82 (30.6)	.67 (44.5)
	50	10%	.86 (17.8)	.78 (16.5)	.84 (19.5)	.75 (25.5)
	50	20%	.85 (18.0)	.62 (15.0)	.73 (17.8)	.63 (31.7)

It is interesting that all the methods did least well for the case of lognormal data, which is probably the case of greatest interest. It was conjectured that the less than optimal coverage might have been due mainly to the significant portion of cases where the Box-Cox method did not recommend the logarithmic transformation.

There is strong support in the literature (see, for example, Hinkley and Runger (1984)) for the idea of doing the analysis assuming a fixed power transformation. Table 3 shows the coverages resulting from fixing the transformation at no transformation or at the logarithmic and applying the results to normal and lognormal data.

It is clear that applying the logarithmic transformation when the data are, in fact, lognormal brings the coverages right up to .89-.90. Hence, it is probably the incorrect Box-Cox choices in the original simulation of Table 2 that are causing the reduced coverages. Applying no transformation when the data are lognormally distributed results in a severe reduction in coverage (.73-.82). Hence, it is clear that estimating the transformation as in Table 2 yields better coverages; knowing the right transformation as in Table 3 does even better.

The upper part of Table 3 is interesting in that we don't seem to pay much of a penalty in coverage for assuming that a lognormal transformation is appropriate for normally distributed data.

Table 3 Simulation results: coverage proportions of delta method confidence intervals (nominal 90% coverage) for estimates of censored means using the indicated transformations.

Correct Transf. to Normality	Sample Size	Censoring	Transformation None	to Normality Used Log
None	20	10%	.89 (.73)	.90 (.73)
	20	20%	.89 (.75)	.89 (.71)
	50	10%	.89 (.47)	.91 (.47)
	50	20%	.89 (.48)	.88 (.45)
log	20	10%	.82 (28.5)	.90 (31.2)
	20	20%	.78 (31.7)	.89 (31.9)
	50	10%	.82 (19.8)	.90 (19.4)
	50	20%	.73 (22.0)	.89 (19.7)

The average biases and mean square errors for the maximum likelihood estimators are given in Table 4. It is interesting to see that on the average, the bias of delta method is only about 1% of the mean.

Table 4 Bias and mean square error simulation results for censored mean estimation using the Box-Cox transformation.

Correct Transf. to Normality	Sample Size	Cens.	Maximum Likelihood	
			Average Bias (400 samples)	Average MSE (400 samples)
None $E(x)=4$	20	10%	.003	.049
	20	20%	.004	.051
	50	10%	.003	.021
	50	20%	.004	.022
Square Root $E(x)=6.5$	20	10%	-.023	.317
	20	20%	-.056	.353
	50	10%	-.014	.135
	50	20%	-.056	.353
log $E(x)=33.115$	20	10%	-.36	34.58
	20	20%	-.28	35.52
	50	10%	-.43	84.47
	50	20%	-.68	87.09

To summarize, the delta method seems to do the best over all alternatives. Since the coverage can be affected by either by choosing the wrong transformation all of the time or by choosing the wrong transformation some of the time, it is probably advisable to run the Box-Cox over a reasonable number of environmental samples from a given category and then use the winning transformation to construct the confidence intervals for all such samples.

A BASIC computer program for PC-compatible microcomputers and a FORTRAN program are available from the Research Division of the California Air Resources Board. Refer to program MNWDL when requesting copies. The programs allow one to determine a power transformation using the data and the Box-Cox method or to specify a transformation. The program always outputs: (1) the estimated mean in the transformed and original scales and (2) its standard error and 90% and 95% delta method confidence intervals in the original scale. Although the percentile bootstrap and bias-corrected percentile bootstrap confidence intervals appeared to give less satisfactory coverages, the user can specify that these intervals be computed as well. Generally, 1000 bootstrap replications will be sufficient. For a sample of size 20 they will take about 10 minutes on a microcomputer equipped with an 8087 math computation chip or about 1 minute on a microcomputer with an 80386 processor. The delta method computations are very fast on any machine. Program documentation and listings are given in Appendix B.

We illustrate the approach on a 24-hour field sample taken during ambient air monitoring of ethyl parathion in the Imperial Valley, California. The data are given in Table 5; five of the fourteen values are below the detection limits. The original investigators computed means

Table 5 Heber Station Ethyl Parathion Concentrations, Imperial Valley, California. (Values in parentheses indicate observed (1) or censored (2)) ($\mu\text{g}/\text{m}^3$)

.010(2)	.010(2)	.010(2)	.010(2)	.018(1)	.032(1)	.012(1)
.015(1)	.010(2)	.078(1)	.092(1)	.023(1)	.018(1)	.010(1)

of this sample in several ways, obtaining: a mean of .033 for the nine samples above the detection limit, a mean of .025 with censored values replaced by .01, and a mean of .023 with censored values replaced by .005. Applying the Box-Cox procedure gave values for the log likelihood of 24.58, 24.29, 23.73 and 21.90 corresponding to the logarithmic, 4th root, and square root transformations and no transformation, respectively; accordingly, the log transformation was used in the analysis. The maximum likelihood estimator is .0233, which is closest to the mean for all samples exceeding .005. The 90% confidence intervals are shown in Table 6

Table 6 90% Confidence Intervals for Ethyl Parathion Means, Imperial Valley, California (Estimated Mean is .0233)

Method	Lower Limit	Upper Limit
Delta	.0101	.0366
Percentile	.0118	.0299
BC Percentile	.0184	.0446

and indicate that the three methods give somewhat different results with the bootstrap method giving shorter intervals. On the basis of the simulations, these intervals probably do not attain a coverage of 90%.

6. Recommendations

This report has concentrated on the development of a procedure using transformation, the EM algorithm, and maximum likelihood estimation to estimate means of small samples of non-normally distributed environmental data which have some observations below the detection limit.

In general, we concluded through the use of simulations that the confidence intervals for the mean were best when the correct transformation is known and less well determined when the transformation had to be estimated. Hence, it seems best to analyze a large number of consistent samples with the hope that they all follow approximately the same probability law; for example, they may all be approximately lognormally distributed.

The simulations also indicated that, over the conditions considered, namely sample sizes of 20 and 50 with 10% and 20% censoring, the approximate delta method intervals based on the large-sample properties of the maximum likelihood estimator did better than bootstrapping. If the transformation was known, the 90% delta method intervals covered the true mean about 90% of the time in all cases whereas the bootstrap coverages were as low as 62% for a nominal 90% interval. If the transformation was not known, the coverages achieved by the delta method were generally at least 85% for a nominal 90% interval. Hence, the delta method is recommended on the basis of this study.

7. Acknowledgements

Dr. John Moore of the Air Resources Board posed the problem considered in this report and contributed to the solution. We are also indebted to the Editor, an Associate Editor and two reviewers for Technometrics who contributed numerous suggestions including an investigation of the properties of the delta method.

References

- Abramowitz, M. and Stegun, I. A. (1964). Handbook of Mathematical functions. NBS Applied Math Series 55. U. S. Govt. Printing Office, Washington, D.C. 20402
- Aitkin, M. (1981). A note on the regression analysis of censored data. Technometrics 23 161-163.
- Blandford, R. R. and Shumway, R. H. (1982). Magnitude-Yield for nuclear explosions with stations containing clipped and low amplitude signals. VSC-TR-12. Seismic Data Analysis Center, Teledyne-Geotech, Alexandria, Virginia 22314.
- Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations (With discussion). J. Royal Statist. Soc. B 39 211-252.
- Cohen, A. C. (1959). Simplified estimators for the normal distribution when the samples are singly censored or truncated. Technometrics 1 217-237.
- Cramer, H. (1946). Mathematical Methods of Statistics. Princeton Univ. Press, Princeton, New Jersey.
- Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. J. Royal Statist. Soc. B 39 1-38.
- Efron, B. (1979). Bootstrap methods: Another look at the jackknife. Ann. Statist. 7 1-26.
- Efron, B. (1982). The jackknife, the bootstrap and other resampling plans. Soc. Ind. Appl. Math. CBMS Natl. Sci. Found. Monogr. 38.
- Efron, B. (1985). Bootstrap confidence intervals for a class of parametric problems. Biometrika 72 45-58.
- Efron, B. (1987). Better bootstrap confidence intervals. (With discussion). J. Amer. Statist. Assoc. 82 171-200.
- Efron, B. and Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Statistical Science 1 54-77.
- Gilliom, R. H. and Helsel, D. R. (1986). Estimation of distributional parameters for censored traced level water quality data 1. and 2. Water Resources Research 22 135-155.
- Gleit, A. (1985). Estimation for small normal data sets with detection limits. Env. Sci. Tech. 19 1206-1213.
- Hall, P. (1986). On the bootstrap and confidence intervals. Ann. Statist 14 1431-1452.

- Hinkley, D. V. and Runger, G. (1984). The analysis of transformed data. (With discussion). J. Amer. Statist. Assoc. 79 302-320.
- Johnson, R. A. and Wichern, D. W. (1988). Applied Multivariate Statistical Analysis 2nd ed. Prentice Hall, Englewood Cliffs.
- Loh, W. (1987). Calibrating confidence coefficients. J. Amer. Statist. Assoc. 82 155-162.
- Poirer, D. J. (1976). The use of the Box-Cox transformation in limited dependent variable models. J. Amer. Statist. Assoc. 73 284-287.
- Ringdal, F. (1975). Maximum likelihood of seismic magnitude. Bull. Seismolog. Soc. Amer. 66 789-802.
- Shumway, R. H., Azari, A. S. and Johnson, P. Estimating mean concentrations under transformation for environmental data with detection limits. Submitted to Technometrics.
- Wu, C. F., (1983). On the convergence properties of the EM algorithm. Ann. Statist. 11 95-103.

APPENDIX A: Technical Details

We summarize here the technical details and equations needed to apply the estimation procedures in the body of the report. A full technical exposition is in Shumway et al (1988).

To set the notation, assume that N independent observations are available of which n_1 are observed and n_2 are below detection limits. Suppose that the sample is denoted by x_1, x_2, \dots, x_N and that if the observation x_i is censored, we know only that $x_i \leq T_i$ where T_i is some lower detection threshold, allowed to differ for each sample value. If x_i is observed, we write $x_i > T_i$ although the value of T_i in this case is irrelevant.

Define the transformed variables

$$y_i = \begin{cases} (x_i^\lambda - 1)/\lambda & \lambda \neq 0 \\ \ln x_i & \lambda = 0 \end{cases} \quad (A1)$$

when x_i is observed and the transformed thresholds

$$T_i^* = \begin{cases} (T_i^\lambda - 1)/\lambda & \lambda \neq 0 \\ \ln T_i & \lambda = 0 \end{cases} \quad (A2)$$

when x_i is censored. This is the Box-Cox transformation (see Box and Cox (1964)) which we consider for x_i positive ($x_i > -1/\lambda$) and $\lambda = 0, \frac{1}{4}, \frac{1}{2}$ and 1.

Now we may write the log likelihood of the original observations x_1, \dots, x_N , assuming the transformed observation y_1, \dots, y_N are normally distributed with mean μ and variance σ^2 , as

$$\begin{aligned} \ln L(\lambda, \mu, \sigma) = & -\frac{n_1}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{x_i > T_i} (y_i - \mu^2) + \sum_{x_i > T_i} J_i(\lambda) \\ & + \sum_{x_i \leq T_i} \Phi(Z_i) \end{aligned} \quad (A3)$$

where the notation $x_i \leq T_i$ denotes summing over the censored values, $J_i(\lambda)$ is the Jacobian

$$J_i(\lambda) = \begin{cases} (\lambda - 1) \ln x_i & \lambda \neq 0 \\ -\ln x_i & \lambda = 0 \end{cases}, \quad (A4)$$

$$Z_i = \frac{T_i^* - \mu}{\sigma} \quad (A5)$$

and

$$\Phi(z) = \int_{-\infty}^z \varphi(x) dx \quad (A6)$$

denotes the normal cumulative distribution function with

$$\varphi(x) = (2\pi)^{-1/2} \exp\left\{-\frac{1}{2} x^2\right\} \quad (A7)$$

The normal probability density function.

Maximizing the log likelihood (B3) will lead to estimators for μ and σ , the mean and standard deviation in the transformed scale. We are interested in the maximum likelihood estimators for the mean of x , namely

$$E(x) = \int_{-\infty}^{\infty} (\lambda y + 1)^{1/\lambda} \varphi\left(\frac{y - \mu}{\sigma}\right) \frac{d\mu}{\sigma} \quad (A8)$$

which is nonlinear function of the parameters λ, μ and σ . When $\lambda=0$, for example, corresponding to the lognormal distribution,

$$E(x) = \exp\left\{\mu + \frac{1}{2} \sigma^2\right\}. \quad (A9)$$

The values of $E(x)$ for $\lambda = \frac{1}{4}, \frac{1}{2}$ and 1 are shown in Table B1.

Table B1 Mean Functions in the Original Scale and Partial Derivatives

λ	$E(x)$	$\frac{\partial E(x)}{\partial \mu}$	$\frac{\partial E(x)}{\partial \sigma}$
0	$\exp\left\{\mu + \frac{1}{2} \sigma^2\right\}$	$\exp\left\{\mu + \frac{1}{2} \sigma^2\right\}$	$\sigma \exp\left\{\mu + \frac{1}{2} \sigma^2\right\}$
.25	$\frac{3}{16} \sigma^2 + \frac{3}{8} \sigma^2 \left(\frac{1}{4}\mu + 1\right)^2 + \left(\frac{1}{4}\mu + 1\right)^4$	$\left(\frac{1}{4}\mu + 1\right)^3 + \frac{3}{16} \sigma^2 \left(\frac{1}{4}\mu + 1\right)$	$\frac{3}{8} \sigma + \frac{3}{4} \left(\frac{1}{4}\mu + 1\right)^2$
.50	$\left(\frac{1}{2}\mu + 1\right)^2 + \frac{1}{4} \sigma^2$	$\left(\frac{1}{2}\mu + 1\right)$	$\frac{1}{2} \sigma$
1.00	$(\mu + 1)$	1	0

The log likelihood function (B3) can be maximized by the Newton-Raphson or EM algorithms to get maximum likelihood estimators for λ, μ and σ and then, by substitution into (B8), the maximum likelihood estimator for $E(x)$.

In this report, we use the EM algorithm of Dempster, Laird and Rubin (1977). The algorithm operates on the log likelihood

$$\ln L'(\lambda, \mu, \sigma) = -\frac{N}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^N (y_i - \mu)^2 + \sum_{i=1}^N J_i(\lambda) \quad (A10)$$

written as though no observations were censored. Iterations are defined as successively maximizing the expectation of the complete-data log likelihood (B10) conditioned on the censoring pattern. For example, if μ_k and σ_k are the current estimators at iteration k , the EM algorithm obtains μ_{k+1} and σ_{k+1} by maximizing

$$\begin{aligned} E_k \left[\ln L'(\lambda, \mu, \sigma) | \text{censoring} \right] = & -\frac{N}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{x_i > T_i} (y_i - \mu)^2 \\ & - \frac{1}{2\sigma^2} \sum_{x_i \leq T_i} E_k \left[(y_i - \mu)^2 | y_i \leq T_i^* \right] + \sum_{i=1}^N J_i(\lambda) \end{aligned} \quad (A11)$$

over μ and σ for a fixed λ . This procedure increases (B3) at each step and converges to the unique maximizer by results obtained by Wu (1983). Scanning the resulting maximizers over λ leads to the final estimator. The updated values for μ and σ at each stage are computed using

$$\hat{\mu}_{k+1} = \frac{1}{N} \left\{ \sum_{x_i > T_i} y_i + \sum_{x_i \leq T_i} E_k(y_i | y_i \leq T_i^*) \right\} \quad (A12)$$

and

$$\sigma_{k+1}^2 = \frac{1}{N} \left\{ \sum_{x_i > T_i} (y_i - \mu_k)^2 + \sum_{x_i \leq T_i} E_k \left[(y_i - \mu_k)^2 | y_i \leq T_i^* \right] \right\} \quad (A13)$$

where the conditional means and variances are computed from

$$E_k \left[(y_i | y_i \leq T_i^*) \right] = \mu_k - \sigma_k \frac{\varphi(Z_i)}{\Phi(Z_i)} \quad (A14)$$

$$E_k \left[(y_i - \mu)^2 | y_i \leq T_i^* \right] = \sigma_k^2 \left[1 - Z_i \frac{\varphi(Z_i)}{\Phi(Z_i)} \right] \quad (A15)$$

where Z_i is defined in (B5).

The above procedure leads to maximum likelihood estimators for λ , μ , and σ and hence, by substituting into (B8), for $E(x)$. We may also develop a confidence interval for $E(x)$ by noting that it is a function of the parameter vector $\underline{\theta} = (\mu, \sigma)'$. It is convenient and realistic to regard λ as being fixed after the arguments in Hinkley and Runger (1984). An estimator for the variance-covariance matrix of $\hat{\underline{\theta}}$ is

$$\text{cov}(\hat{\underline{\theta}}) = \left\{ - \frac{\partial^2 \log L}{\partial \underline{\theta} \partial \underline{\theta}'} \right\}_{\hat{\underline{\theta}}}^{-1} \quad (\text{A16})$$

where $\log L$ is abbreviated for (B3). The elements of

$$\frac{\partial^2 \log L}{\partial \underline{\theta} \partial \underline{\theta}'} = \begin{pmatrix} L_{\mu\mu} & L_{\mu\sigma} \\ L_{\sigma\mu} & L_{\sigma\sigma} \end{pmatrix} \quad (\text{A17})$$

are computed as

$$L_{\mu\mu} = - \frac{n_1}{\sigma^2} - \frac{1}{\sigma^2} \sum_{x_i \leq T_i} (Z_i R_i + R_i^2), \quad (\text{A18})$$

$$L_{\mu\sigma} = - \frac{2}{\sigma^3} \sum_{x_i > T_i} (y_i - \mu) - \frac{1}{\sigma^2} \sum_{x_i \leq T_i} (Z_i R_i^2 + Z_i^2 R_i - R_i) \quad (\text{A19})$$

and

$$L_{\sigma\sigma} = \frac{n_1}{\sigma^2} - \frac{3}{\sigma^4} \sum_{x_i > T_i} (y_i - \mu_k)^2 - \frac{1}{\sigma^2} \sum_{x_i \leq T_i} (Z_i^2 R_i^2 + Z_i^3 R_i - 2Z_i R_i) \quad (\text{A20})$$

where

$$R_i = \frac{\varphi(Z_i)}{\Phi(Z_i)} \quad (\text{A21})$$

The basis for the delta method of Cramer (1946) is that $\hat{E}(x)$ is a function of $\hat{\underline{\theta}} = (\hat{\mu}, \hat{\sigma})'$ which are expected to be jointly asymptotically

normal so that $\text{var}(\hat{E}(x))$ can be consistently estimated by

$$\text{var}(\hat{E}(x)) = \left(\frac{\partial E(x)}{\partial \underline{\theta}} \right)'_{\hat{\underline{\theta}}} \hat{\text{cov}}(\hat{\underline{\theta}}) \left(\frac{\partial E(x)}{\partial \underline{\theta}} \right)_{\hat{\underline{\theta}}} \quad (\text{A22})$$

where the 2×1 vectors of partial derivatives of $E(x)$ are as given in Table B1. This implies an approximate $100(1-\alpha)$ percent confidence interval of the form

$$\hat{E}(x) \pm Z_{\alpha/2} \sqrt{\hat{\text{cov}}(\hat{E}(x))}. \quad (\text{A23})$$

where Z_{α} is the $100(1-\alpha)$ percentile of the standard normal.

